

Coarse-Grained Model of Collagen Molecules Using an Extended MARTINI Force Field

Alfonso Gautieri,^{†,‡} Antonio Russo,[†] Simone Vesentini,[†] Alberto Redaelli,[†] and Markus J. Buehler^{*,‡,§}

Biomechanics Group, Department of Bioengineering, Politecnico di Milano, Via Golgi 39, 20133 Milan, Italy, Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Room 1-235A&B, Cambridge, Massachusetts, and Center for Computational Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, Massachusetts

Received January 9, 2010

Abstract: Collagen is the most abundant protein in the human body, providing mechanical stability, elasticity, and strength to connective tissues such as tendons, ligaments, and bone. Here, we report an extension of the MARTINI coarse-grained force field, originally developed for lipids, proteins, and carbohydrates, used to describe the structural and mechanical properties of collagen molecules. We identify MARTINI force field parameters to describe hydroxyproline amino acid residues and for the triple helical conformational structure found in collagen. We validate the extended MARTINI model through direct molecular dynamics simulations of Young's modulus of a short 8-nm-long collagen-like molecule, resulting in a value of approximately 4 GPa, in good agreement with earlier full atomistic simulations in explicit solvent as well as experimental results. We also apply the extended MARTINI model to simulate a 300-nm-long human type I collagen molecule with the actual amino acid sequence and calculate its persistence length from molecular dynamics trajectories. We obtain a value of 51.5 ± 6.7 nm for the persistence length, which is within the range of earlier experimental results. Our work extends the applicability of molecular models of collagenous tissues by providing a modeling tool to study collagen molecules and fibrils at much larger scales than accessible to existing full atomistic models, while incorporating key chemical and mechanical features and thereby presenting a powerful approach to computational materiomics.

1. Introduction

Collagen (or tropocollagen) molecules represent the most abundant protein building block in the human body, where they provide primarily mechanical stability, elasticity, and strength to connective tissues such as tendons, ligaments,

and bone.¹ Collagen's primary structure is formed by a sequence of triplets (glycine-X-Y)_n, where X and Y can represent any amino acids, but mostly proline and hydroxyproline. In particular, the presence of glycine every three amino acids guarantees the stability of the tertiary molecular structure, characterized by a right-handed triple helix² with an overall length of ~ 300 nm and diameter of ~ 1.5 nm, leading to a great aspect ratio of ~ 200 .

Collections of collagen molecules form well-defined hierarchical structures in tissues that give rise to fibrils and fibers, which universally represent the basis of most connective tissues.³ Collagen provides mechanical integrity to connective tissues through, above all, a great resistance to

* Corresponding author phone: +1-617-452-2750; e-mail mbuehler@MIT.EDU.

[†] Politecnico di Milano.

[‡] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

[§] Center for Computational Engineering, Massachusetts Institute of Technology.

stretching along the main direction of molecules, fibrils, and fibers. Moreover, collagen is capable of providing outstanding mechanical features as it interacts with other biological molecules and minerals, such as glycosaminoglycans or hydroxyapatite in bone. For example, in bone tissue, the toughness is much greater than that of collagen or hydroxyapatite alone. This is due to the fact that collagen, with its capacity to yield and dissipate energy, contributes to the material's enhanced toughness by providing an effective means to confer shear stresses between hydroxyapatite crystals and to dissipate mechanical energy through molecular unfolding and intermolecular shear.⁴ The previous examples underline that, physiologically, collagen cannot solely be considered as a single molecule but must be studied in the context of functional networks that are formed on the basis of a large number of collagen molecules and other biopolymers. The significance of collagen in controlling key material properties in connective tissues is further evident from mutations in collagen, which result in incorrect protein folding that causes several severe pathologies, such as Ehlers–Danlos syndrome, Alport syndrome,⁵ or *osteogenesis imperfecta*.⁶

Great efforts have been put forth in recent years focused on characterizing the mechanical properties of collagen molecules and fibrils using both experimental and computational (theoretical) approaches. Some of the earliest experiments analyzed type I collagen molecules in terms of their stiffness. For example, Sasaki and Odajima⁷ estimated the Young's modulus of collagen molecules by using X-ray diffraction techniques. Bozec and Horton⁸ used an AFM setup to evaluate both the topographical and mechanical properties of the collagen molecules. Optical trapping methods have also been used for the assessment of collagen's mechanical properties, as reported by Sun and co-workers.⁹ A review of the global efforts in the understanding of the structure and properties of collagen-based tissues are nicely illustrated in recent works.¹⁰ Nonetheless, a persistent limitation of experimental analyses is the lack of details about nanoscale phenomena, and limitations in sample preparation. To complement experimental approaches, molecular simulation provides an alternative approach to describe the molecular mechanics of collagen from a bottom-up perspective. Most earlier studies, often based on collagen-like peptides obtained from X-ray crystallography,¹¹ focused on relatively short collagen molecules^{12–17} limited to ~ 10 nm length. However, the contour length of collagen molecules is ~ 300 nm, which represents a scale that is not yet accessible to full atomistic molecular modeling. Further, the study of collagen fibrils and fibers is currently not feasible with full atomistic simulation, which represents an important limitation since this scale in the structural hierarchy of collagen is most relevant for physiological function. Other limitations exist with respect to the accessible time scale, where most full-atomistic simulations are limited by a few hundred nanoseconds. However, many relevant materials phenomena such as tissue deformation and failure (e.g., under disease conditions) emerge at much longer time scales.

A promising strategy to overcome these limitations is to decrease the number of degrees of freedom by grouping

atoms into pseudoatoms (or particles) referred to as beads. This represents the basis of the so-called coarse-grained approach,^{18,19} where, starting at the nanoscale, it is possible to derive parameters for higher hierarchical levels, up to the macroscale by systematically feeding information from smaller, more accurate to larger, more coarse levels. The development of a coarse-grained model provides a powerful path toward reliable modeling of full-length collagen molecules and its higher-level hierarchical structures. Specifically, coarse-grained models allow the study of more complex systems, up to micrometer dimension and millisecond duration.²⁰ Earlier work of coarse-graining collagen molecules grouped hundreds of atoms into particles or beads.^{14,15} This level of coarse-graining, however, is at a relatively coarse level where information about biochemical features (e.g., amino acid sequence) cannot be represented directly. However, the incorporation of biochemical features is crucial in a computational materiomics approach where material properties are elucidated at multiple scales, including the level of amino acid sequence that links to genetic level information.

Several other coarse-grained models suitable for proteins have been developed and successfully applied in earlier works.²¹ Particularly, the MARTINI coarse-grained model,²² developed by Marrink and co-workers, was initially applied to membrane lipids, later extended to proteins,²³ and recently also to carbohydrates.²⁴ It has been successfully applied to gain insights into different biological molecules such as membrane proteins,^{25,26} ion channels,²⁷ and liposomes.²⁸

The MARTINI model provides a suitable level of coarse-graining, as it retains information about the chemistry specific to the amino acid sequence (as side chains are modeled depending on the type of the amino acids). All amino acids in the MARTINI force field are modeled with a number of beads that varies depending on the steric volume of each amino acid.²³ The general mapping rule is that four heavy atoms (that is, non-hydrogen atoms) are grouped together into one bead. Further, one bead describes the backbone, while others are added to represent the side chain. The number of beads used to model a specific residue therefore varies depending on the dimensions of the side chain of the amino acid. Small amino acids such as glycine or alanine are described by just one bead, while larger amino acids, like phenylalanine, tyrosine, and tryptophan are modeled with up to five beads. The model also takes into account the polarity of every bead, described by a letter (P, polar; C, apolar; N, nonpolar; Q, charged) and a number (from 1, low polarity, to 5, high polarity). Further, a letter is used to characterize a residue's hydrogen bonding capability (d, donor; a, acceptor; da, donor and acceptor; 0, none). These bead types, which correspond to atom types in the atomistic modeling framework, are used to describe the nonbonded interactions between beads. For each pair of bead types, a set of parameters for electrostatic and van der Waals potentials is defined. The best choice of particle types for amino acids was obtained by the authors of the original MARTINI force field for proteins by comparing simulation results and experimental measurements of water/oil partitioning coefficients of the amino acid side-chain analogues.²³

However, although the MARTINI force field is suitable for a wide number of applications and its validity has been proven through several simulations and comparisons both with atomistic and with experimental data, it cannot be applied to model collagen molecules in the presently available formulation. This is because of two reasons. First, the existing formulations of MARTINI lack parameters for hydroxyproline (a nonstandard amino acid, found solely in collagen and formed through post-translational modification of proline). Second, it lacks parameters to describe the triple helical configuration of a collagen molecule. The work reported here overcomes these limitations and addresses two major points. First, we extend the MARTINI model to enable the modeling of collagen molecules by including parameters for hydroxyproline and by adding parameters to describe the collagen triple helix. Second, the extended MARTINI force field is applied to perform coarse-grained molecular dynamics simulations of a full-length model of a collagen type I molecule, used to calculate mechanical parameters such as Young's modulus and the persistence length, which allows us to validate the new model through a comparison with earlier computational and experimental results.

2. Materials and Methods

We use a computational multiscale approach to generate an atomistic-informed coarse-grained model of tropocollagen in the framework of the MARTINI force field. The parameters of the coarse-grained model are obtained through a combination of experimental and full-atomistic modeling data to identify the parameters for hydroxyproline residues, as well as through a statistical analysis of collagen-like PDB entries to assess the geometrical features of the coarse-grained tropocollagen triple helix. The results of the parameterization are fed into the coarse-grained tropocollagen model in the spirit of a multiscale simulation approach.

All full atomistic simulations are carried out using the GROMACS code^{29,30} and the GROMOS96 43a1 force field, which includes also parameters for the hydroxyproline residue. The protein molecules are entirely solvated in periodic water boxes with single point charge (SPC) water molecules as the solvent model. SETTLE (for water) and LINCS algorithms are used to constrain covalent bond lengths involving hydrogen atoms, thus allowing a time step of 2 fs. Nonbonding interactions are computed using a cutoff for the neighbor list at 1 nm, with a switching function between 0.8 and 0.9 nm for van der Waals interactions, while the Particle-Mesh Ewald sums (PME) method is applied to describe electrostatic interactions. In the case of charged peptides, counterions (Cl^- or Na^+) are added in order to keep the system neutral. The preliminary system energy minimizations are performed by using a steepest descent algorithm until convergence. The systems are then equilibrated through *NPT* molecular dynamics simulations at a temperature of 310 K (37 °C). In order to assess the elastic constants of the coarse-grained model, simple atomistic oligopeptides are considered. A pulling force is applied along the molecular axis, where one molecular end is held fixed using a rigid constraint while the other is pulled through the use of a virtual spring with a known elastic constant. This setup

corresponds to the steered molecular dynamics setup used in similar atomistic works.^{13,31} In particular, the value of the spring elastic constant is chosen to be $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$, while the deformation rate is set at 0.1 m/s. The steered molecular dynamics simulations performed on the atomistic systems provide the reference force-extension behaviors.

Coarse-grained molecular dynamics simulations are carried out using the GROMACS code and the MARTINI force field with the inclusion of the parameters found in this work for tropocollagen. The models are entirely solvated in periodic water boxes using the coarse-grained water model provided within the MARTINI force field. In the case of charged molecules, counterions (Cl^- or Na^+) are added in order to keep the system neutral. The preliminary system energy minimizations are performed by using a steepest descent algorithm until convergence. The systems are then equilibrated through *NPT* molecular dynamics simulations at a temperature of 310 K (37 °C) using a time step of 20 fs. Steered molecular dynamics simulations are performed using the same setup as described for atomistic simulations. Finally, the whole structure of the heterotrimeric type I collagen molecule is studied. The primary structure is obtained from PubMed (entry number NP_000079 for α_1 chain and NP_000080 for α_2 chain), and the atomistic triple helical structure is built using the software THeBuScr³² (Triple-Helical collagen Building Script). The atomistic representation is then coarse-grained and divided into five 60-nm-long segments, which are energy minimized and simulated in explicit solvent at finite temperature for 350 ns each.

3. Results

3.1. Extension of the MARTINI Force Field to Include Hydroxyproline. The existing MARTINI protein force field represents all 20 naturally occurring amino acids but lacks hydroxyproline, an amino acid produced in collagen synthesis via post-translational modification of proline (Figure 1A). In order to extend the MARTINI model for the study of collagen molecules, hydroxyproline parameters must be introduced. In the MARTINI model, proline is modeled through the use of two beads, one for the backbone (bead type C5) and one for the side-chain (bead type AC2), as shown in Figure 1A. Hydroxyproline derives directly from proline via the addition of a hydroxyl group on its side-chain. Therefore, hydroxyproline is also modeled using two beads in the extended MARTINI model. For the backbone bead, we maintain the same bead type as for proline. The side chain, however, due to the presence of the hydroxyl group, shows a higher polarity level than proline. This aspect was already demonstrated in the work of Black and Mould,³³ who calculated an index that takes into account the hydrophobicity of all amino acid side-chains, including hydroxyproline. Matching the bead types assigned by Marrink and the hydrophobicity values derived from Black and Mould, the hydroxyproline side-chain bead polarity, in the MARTINI notation, is determined. This parameter ranges from 0 to 1, where 1 is the most hydrophobic amino acid (phenylalanine), while 0 is the most hydrophilic one (arginine). The value reported for hydroxyproline is 0.527, which is found between

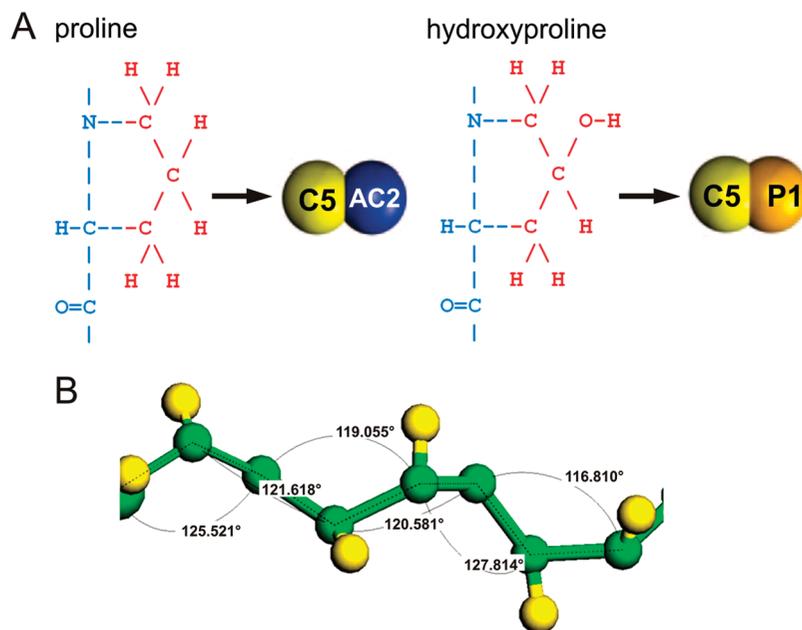


Figure 1. Parameterization of the force field of the MARTINI force field including the hydroxyproline residue (panel A) and geometrical features (angles) of the coarse-grained collagen triple helix (panel B). Hydroxyproline directly derives from proline, adding a hydroxyl group to its side chain, giving rise to an increased polarity level. Atomistic structures of both proline and hydroxyproline are shown on the left side of panel A, while the bead representation using MARTINI model notation is shown on the right side of panel A. Panel B shows the analysis of the angle between backbone beads of collagen-like peptides. The statistical analysis is performed on five collagen PDB entry (2DRT, 1K6F, 1QSU, 1CAG, 1V6Q). These collagen-like peptides are coarse-grained, and the bond lengths, angles, and dihedrals between backbone beads are analyzed.

the threonine (0.450) and cysteine (0.680) values. In the MARTINI force field, the threonine side chain is modeled as a P1 bead, while cysteine is modeled as a C5 bead. Considering this result, hydroxyproline is assigned a P1 value for the hydroxyproline side-chain, since its hydrophobicity value is closer to the value for threonine.

3.2. Extension of the MARTINI Force Field to Describe the Triple Helical Structure of Collagen Molecules. The MARTINI model also takes into account the secondary structure of a protein, such as the α -helix, β -sheet, extended turns, or bends. The characteristic triple helical structure of collagen molecules cannot be described by any of the secondary structure parameters currently included in the MARTINI model. Therefore, it is necessary to determine additional parameters for the triple helical collagen structure. To achieve this, we assess the bond distances, angles, and dihedral parameters specific to the collagen triple helix. The MARTINI bond potential forms are maintained in our extended formulations, which are described as follows:

$$V_b = \frac{1}{2}K_b(d_{ij} - d_b)^2 \quad (1)$$

$$V_a = \frac{1}{2}K_a[\cos(\varphi_{ijk}) - \cos(\varphi_a)]^2 \quad (2)$$

$$V_d = K_d[1 + \cos(n\psi_{ijkl} - \psi_d)] \quad (3)$$

where V_b , V_a , and V_d are the potential energy terms of bond stretching, angle deformation, and dihedral deformations, respectively. Bonded interactions act between bonded sites i , j , k , and l with a distance at the equilibrium of d_b , angle

φ_a , and dihedral angles Ψ_d and with elastic stiffness of K_b , K_a , and K_d (for bond, angle, and dihedral, respectively).

The triple helical geometrical parameters of the collagen triple helix (that is, d_b , φ_a , and Ψ_d) are obtained by performing a statistical analysis on a set of five collagen-like molecules available in the Protein Data Bank³⁴ (identification codes: 2DRT, 1K6F, 1QSU, 1CAG, and 1V6Q). The bond lengths between backbone beads, bonding angles, and dihedral angles are computed on the basis of these five crystallographic structures. From a statistical analysis of different collagen molecules, we determine a bond reference length (d_b) of 0.365 ± 0.07 nm, a bonding reference angle (φ_a) of $119.2 \pm 8.72^\circ$, and a dihedral reference angle (Ψ_d) of $-89.3 \pm 9.76^\circ$. Figure 1B shows the details of a coarse-grained model of a collagen chain with the angular geometrical parameters.

In order to assess the stiffnesses of bonds, angles, and dihedrals (respectively, K_b , K_a , and K_d), simple atomistic oligopeptides are considered, as shown in Figure 2A. In order to derive the bonding constant K_b , a glycine–proline structure is chosen. In order to derive the angle elastic constant K_a , a glycine–proline–hydroxyproline oligopeptide is considered. Finally, to obtain the dihedral elastic constants K_d , a glycine–proline–hydroxyproline–glycine molecule is studied. These structures are considered both in their atomistic form and in their coarse-grained form, and the parameters of the coarse-grain model (i.e., K_b , K_a , and K_d) are identified matching force–extension curves obtained from atomistic and coarse-grained simulations.

Steered molecular dynamics simulations are performed on the three atomistic systems to obtain reference force–extension behaviors. For the coarse-grained models, the

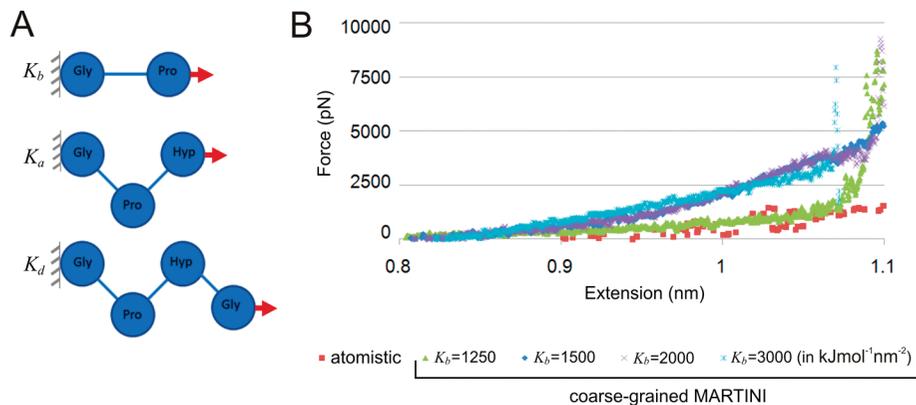


Figure 2. Schematic of the setup used to calculate elastic constants for the coarse-grain force field (panel A) and force-extension plot of atomistic and coarse-grain peptides used to find the optimal value of bonding elastic constant (panel B), varying K_b from 1250 $\text{kJ mol}^{-1} \text{nm}^{-2}$ up to 3000 $\text{kJ mol}^{-1} \text{nm}^{-2}$. The bonding elastic constant (K_b) is obtained by pulling a glycine–proline peptide, the angle elastic constant (K_a) from a glycine–proline–hydroxyproline peptide, and the dihedral elastic constant (K_d) from a glycine–proline–hydroxyproline–glycine peptide. The molecules are considered both in coarse-grained and in atomistic forms and are subjected to steered molecular dynamics simulations. The values of the bond, angle, and dihedral elastic constant are then optimized in order to match the behavior of the atomistic model.

Table 1. Backbone Bonded Parameters for the Collagen Triple Helix (Present Work) and for the Other Secondary Structures, as Included in the Original MARTINI Force Field²³

backbone	d_b [nm]	K_b [$\text{kJ mol}^{-1} \text{nm}^{-2}$]	φ_a [deg]	K_a [kJ mol^{-1}]	Ψ_d [deg]	K_d [kJ mol^{-1}]
collagen triple helix [present work]	0.365	1250	119.2	150	-89.3	100
α -helix	0.35	1250	96	700	60	400
coil	0.35	200	127	25		
extended	0.35	1250	134	25	180	10
turn	0.35	500	100	25		
bend	0.35	400	130	25		

spring constants (that is, the parameters K_b , K_a , and K_d) are initially set equal to the values adopted in the original MARTINI force field for an α helix secondary structure²³ and then gradually changed in order to best fit the atomistic force-extension curves so that they show the same stiffnesses. The procedure is first performed on the dipeptide (glycine–proline) in order to find the optimal value of K_b . Once this parameter is set, we consider a glycine–proline–hydroxyproline peptide in order to find the value of K_a that best approximates the behavior of the corresponding atomistic system. Finally, having fixed K_b and K_a , the glycine–proline–hydroxyproline–glycine peptide is modeled in order to determine the value of K_d (see Figure 2A). Figure 2B shows a plot of force versus deformation in the case of the two amino acid peptides, used to calculate the bonding elastic constant K_b between two beads. The values that best replicate the atomistic behavior are $K_b = 1250 \text{ kJ mol}^{-1} \text{ nm}^{-2}$, $K_a = 150 \text{ kJ mol}^{-1}$, and $K_d = 100 \text{ kJ mol}^{-1}$. Table 1 shows an overview of all model parameters.

Due to coupling of bonded and nonbonded force field terms, the equilibrium geometrical features may in general differ from force field reference values. However, to address this issue, we monitor the equilibrium values and find no significant difference with respect to the reference values.

3.3. Validation of the Extended MARTINI Model: Young's Modulus and Persistence Length of the Collagen Molecule. In order to validate the extended MARTINI force field model, an 8-nm-long collagen-like

molecule (similar to the PDB entry used to assess the geometrical features) is pulled under the same conditions previously described. The chosen molecule is [(glycine–proline–hydroxyproline)₁₀]₃, which represents an “ideal” reference collagen molecule that was used in several earlier molecular dynamics works.^{14,17} The RMSD of the coarse grain peptide during equilibration is compared with the equivalent atomistic counterpart, showing that the models reach similar and stable configurations (see Figure 3A). Figure 3B shows a snapshot of both atomistic and coarse-

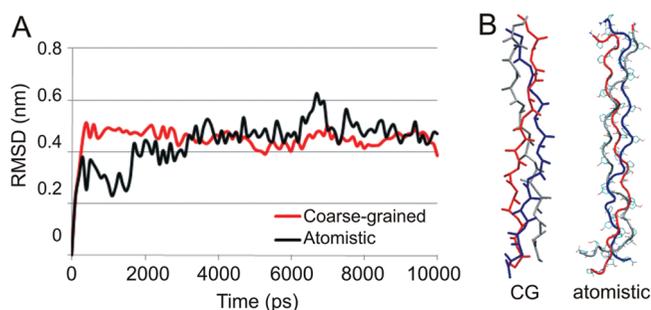


Figure 3. Comparison of the atomistic and coarse-grained structures of a collagen molecule. Panel A displays the root mean squared displacement (RMSD) with respect to the starting structure during 10 ns of simulation time. Panel B shows snapshots of stable collagen triple helical structures in the coarse-grained (CG, left) and atomistic (right) representations. The images show that the triple helical structure is well-maintained in the CG representation.

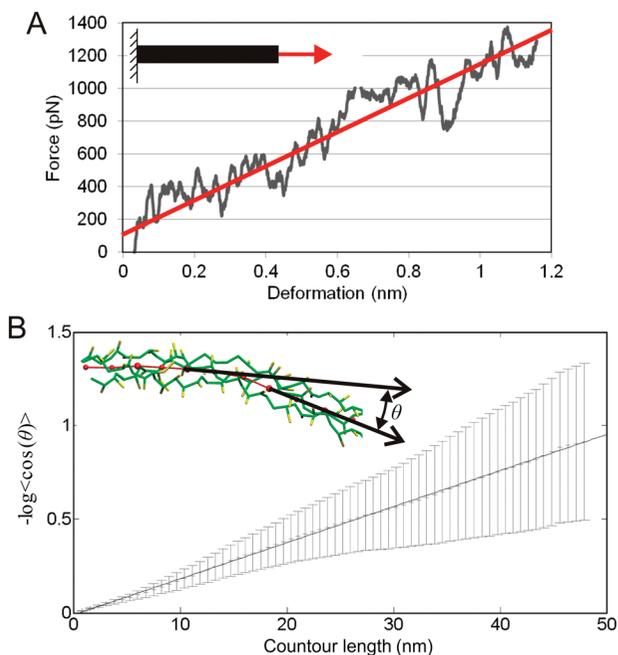


Figure 4. Mechanical analysis of collagen molecules. Panel A: Force-deformation plot up to 15% strain (1.2 nm) for the pulling test of a [(glycine–proline–hydroxyproline)₁₀]₃ collagen-like peptide (straight line) and linear regression (dashed line) used to obtain the molecular elastic constant. Panel B: Plot of $-\log \langle \cos(\theta) \rangle$ versus contour length calculated for the coarse-grain model of the first strand (≈ 60 nm) of the human collagen type I molecule. The linear regression of the data for the five different strands leads to an average persistence length of ~ 51.5 nm for human type I collagen. The inset in panel B shows a schematic of the data analysis approach used to determine the collagen persistence length. The molecule is divided into segments (red lines) which are delimited by the center of mass (red balls) of the three equivalent glycine residues on the three collagen chains. The angle $\theta(s)$ is the angle between two segments that are separated by a contour length s (where s is the sum of the segments' lengths). The value of the angle $\theta(s)$ is obtained averaging over the 350 ns trajectory.

grained collagen peptides. The equilibrated coarse-grained collagen peptide is pulled using a deformation rate of 1 m/s. The Young's modulus of the coarse-grain collagen molecule is then calculated from the slope of the force-deformation plot, as described in previous works.^{17,31} Force-extension plots similar to one performed on a coarse-grained molecule are shown in Figure 4A, together with a linear regression analysis used to obtain the stiffness of the collagen molecule (1052.78 ± 51.23 pN/nm) up to 15% strain. Assuming a cylindrical geometry, and considering a collagen molecular radius of 0.8 nm, the Young's modulus is obtained equal to 4.62 ± 0.41 GPa. Table 2 shows the values of Young's modulus obtained through different approaches, showing that our result agrees well with earlier findings.

Finally, the whole structure of the heterotrimeric type I collagen molecule is investigated in order to derive the persistence length. The full-length coarse-grained tropocollagen model is divided into five 60 nm peptide strands to reduce the computational cost. Then, the flexibility of the structures is analyzed from molecular dynamics trajectories

of 350 ns each, obtaining the persistence length through the following expression:³⁵

$$\log \langle \cos \theta \rangle = -\frac{s}{L_p} \quad (4)$$

As shown in the inset in Figure 4B, the variable θ denotes the angle between two segments along the molecule separated by contour length s (that is, the sum of the segment lengths), and L_p is the persistence length, which provides information about a molecule's flexibility. In order to evaluate the expression defined in eq 4, the molecule is divided into several smaller segments. Each segment (as shown in the inset in Figure 4B), is delimited by the center of mass of the three equivalent glycine residues belonging to the three different collagen chains. The plot of $-\log \langle \cos(\theta) \rangle$ versus s is obtained, where the average of $\cos(\theta)$ is calculated over the full molecular dynamics simulations trajectories. Figure 4B shows the plot of $-\log \langle \cos(\theta) \rangle$ versus the contour length s . The value of the persistence length L_p is then obtained from the inverse of the slope after a linear regression of all data, giving a value of $L_p = 51.5 \pm 0.6.7$ nm.

The persistence length found based on the coarse-grained model is close to the value obtained by Hofmann and co-workers,³⁶ who found a value of 57 ± 5 nm for collagen type I molecule through an electron microscopy analysis. On the other hand, a literature analysis shows that great variability in persistence length value can be found considering different kinds of experimental tests. The pioneering work in this analysis is represented by Utiyama and co-workers,³⁷ who considered sedimentation constants and the intrinsic viscosity of purified collagen molecules and calculated a value for the persistence length close to 130 nm. Saito and co-workers,³⁸ considering the hydrodynamic properties of collagen, derived the intrinsic viscosity and the sedimentation constant of collagen and, from these values, the persistence length, equal to 160–180 nm. In another experiment, Nestler and co-workers³⁹ measured the dynamic viscoelastic properties of dilute solutions of collagen molecules. From the obtained values of intrinsic viscosity and rotational relaxation time, they found a value for the persistence length of about 170 nm. In more recent work, Sun and co-workers⁹ used optical tweezers in order to obtain, from force-extension plots fitted to the Marko–Siggia entropic elastic model,⁴⁰ the collagen persistence length, which was found to be 14.5 \pm

Table 2. Comparison of Young's Modulus of the Single Collagen Molecule Calculated Using Different Experimental and Computational Analysis

type of analysis	Young's modulus (GPa)
X-ray diffraction ⁷	~ 3
Brillouin light scattering ⁴⁵	~ 5.1
estimate based on persistence length ⁹	0.35–12
estimate based on persistence length ³⁶	~ 3
single molecule stretching–atomistic modeling ¹³	~ 4.8
single molecule stretching–reactive atomistic modeling ¹⁴	~ 7
single molecule stretching–atomistic modeling ¹²	~ 2.4

Table 3. Comparison of the Persistence Length (L_p) of Several Biopolymers

polymer	configuration	L_p
titin ⁴⁶	linear protein with two domains	~15 nm
spectrin ⁴⁷	double-strand filament	10–20 nm
collagen [optical tweezers] ⁴⁸	triple helix	14.5 ± 7.3 nm
collagen [MD bending simulations] ¹⁶	triple helix	23.4 nm
collagen [present work, coarse-grain model]	triple helix	51.5 ± 6.7 nm
collagen [electron microscopy] ³	triple helix	57 ± 5 nm
collagen [hydrodynamic properties] ^{37,39,49}	triple helix	130–180 nm
DNA ⁵⁰	double helix	~50 nm
F-actin ³⁵	Filament	0.5–15 μm
intermediate filaments ⁵¹	32 strands filament	~1 μm
microtubules ⁵²	13 protofilaments	0.1–10 mm

7.3 nm. Finally, earlier molecular simulation studies were used to perform bending tests on a collagen-like peptide, predicting the bending stiffness of the collagen molecule and, from this result, evaluate the persistence length of collagen to be in the range of 16–24 nm.¹⁶ The variability that can be found in the literature for the collagen persistence length value depends on the used experimental setup, for example, electron micrographs or optical tweezers, on the experimental conditions (how the sample is extracted and prepared), on the method applied and the parameters used to fit experimental data with theoretical models. Despite the large variability of experimental values, the value obtained in the present work is in good agreement with the consideration of collagen as a semiflexible molecule with a rodlike configuration. A comparison with other biomolecules shows that the persistence length of biomolecules spans several order of magnitudes (see Table 3) and that, as for collagen, in some other cases, a very precise value cannot be established, but rather a range. In the case of microtubules, Pampaloni et al.⁴¹ found that the persistence length value ranges from a few hundred micrometers up to a few millimeters. Soictin shows a certain variability depending on the kind of setup used to derive the persistence length, with values from 0.5 μm up to 15 μm.⁴² Table 3 shows the values of the persistence length of collagen molecules obtained through different approaches, showing that our result agrees reasonably well with earlier findings, albeit tending to be larger than what has been computed on the basis of earlier full atomistic modeling.

The calculation of L_p provides an alternative way to estimate the Young's modulus Y , since

$$Y = \frac{4L_p k_B T}{\pi r^4} \quad (5)$$

where L_p is the persistence length, k_B is Boltzmann's constant, T is the absolute temperature, r is the molecular radius, and Y is the Young's modulus. Considering a temperature of 300 K, $r = 0.8$ nm, and applying eq 5, the Young's modulus Y results to be 0.66 ± 0.08 GPa. The deviation between direct stretching and the estimation of Young's modulus via the persistence length data suggests that eq 5 may not hold, which may reflect the fact that a collagen molecule is not a

structurally homogeneous molecule as assumed in the underlying continuum theory.

4. Discussion and Conclusion

The most important contribution of this article is the extension of the MARTINI coarse-grained force field to include parameters that allow the modeling of collagen molecules. The original MARTINI protein force field lacks parameters for hydroxyproline, which is a nonstandard amino acid but is found frequently in natural collagen molecules. Furthermore, the MARTINI force field can describe a variety of protein secondary structures but lacks some peculiar cases of secondary structures, such as the collagen triple helix, in its original form. These limitations are overcome in the extended MARTINI coarse-grained force field reported here, now enabling the application of coarse-grained MARTINI models to model collagenous protein materials. Due to the broad significance of collagenous tissues in biomechanics, biochemistry, and biology in general, the new model could find wide applicability in many future studies.

Several validation computations have been carried out to ensure that the extended MARTINI model can accurately describe key mechanical and biophysical parameters of collagen molecules. First, we considered the Young's modulus of a collagen molecule by simulating a short collagen-like peptide with the sequence [(glycine–proline–hydroxyproline)₁₀]₃. The collagen molecule was subjected to axial load by using steered molecular dynamics, and the study resulted in a Young's modulus value of 4.62 GPa, which is in good agreement with those obtained both through atomistic setups^{12,13,15,43} and experimental analysis^{44,7,8} (see Table 2). Second, we have computed the persistence length of collagen molecules, leading to $L_p = 51.5 \pm 0.67$ nm. We find reasonable agreement with results from earlier studies (see Table 3).

Due to its contour length of 300 nm, all-atom simulations with explicit solvent are prohibitive since they would require excessive computational resources due to the very large number of particles. The reduction by roughly a factor of 10 in the coarse-grained description provides a significant speedup that facilitates the direct simulation of much longer molecules. Furthermore, the typical time step used in classical molecular dynamics simulations (1–2 fs) allows the modeling only on the nanosecond time scale. However, the coarse-grained model enables one to use much longer time-scales on the order of 20–40 fs. Considering the combined effect of the reduction of the number of particles and the increased time step, the coarse-grained approach leads to a total speedup of 200–400 with respect to atomistic simulations.

With this significant computational speedup, the modeling framework reported here opens many possibilities for future studies, particularly at the scale of collagen fibrils and possibly fibers. We note that the MARTINI coarse grain approach is only valid when the phenomena under study do not involve changes of the secondary structure. In the context of collagen, only events that do not involve unfolding of the collagen triple helix can be correctly modeled using the MARTINI force field formulation presented here. While unfolding of molecules is likely to play an important role in

large deformation and the fracture of collagenous tissues, the mechanical stresses experienced by collagen tissues under small mechanical loads may not lead to unfolding at the molecular level. This is because strain is distributed over several hierarchical levels (i.e., fibers, fibrils, molecules) and involves several concurrent mechanisms (fiber uncrimping, proteoglycan-mediated fibril sliding, molecular slippage, and molecular elongation).

The new method could be particularly useful to predict and analyze the structure of collagen fibrils and even collagen fibers. For example, the coarse-grain approach may facilitate the computational investigation of how changes in the sequence would influence the packing of collagen fibril, helping the understanding of the mechanisms underlying collagen-related diseases, such as *osteogenesis imperfecta*. This, together with the study of the interaction with other relevant biomolecules such as proteoglycans, will provide useful details for the understanding of structure–property relationships in the broader class of collagen materials and as such makes an important contribution to materiomics.

Acknowledgment. M.J.B. acknowledges support through a National Science Foundation CAREER award (CMMI-0642545). This research was supported in part through the MIT-Italy program (Progetto Rocca).

Supporting Information Available: A script to convert an atomistic PDB file to a coarse-grained PDB file (atom2cg.awk) as well as a script to generate MARTINI coarse-grained topologies (seq2itp.pl) are included. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Kadler, K. E.; Baldock, C.; Bella, J.; Boot-Handford, R. P. Collagens at a glance. *J. Cell Sci.* **2007**, *120*, 1955.
- (2) Hyde, T. J.; Bryan, M. A.; Brodsky, B.; Baum, J. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J. Biol. Chem.* **2006**, *281*, 36937–36943.
- (3) Hoffmann, D.; Voss, T.; Kuhn, K.; Engel, J. Localization of flexible sites in thread-like molecules from electron micrographs. Comparison of interstitial, basement membrane and intima collagens. *J. Mol. Biol.* **1984**, *172*.
- (4) Cowin, S. C. The mechanical and stress adaptive properties of bone. *Ann. Biomed. Eng.* **1983**, *11*, 263–95.
- (5) Srinivasan, M.; Uzel, S. G. M.; Gautieri, A.; Keten, S.; Buehler, M. J. Alport Syndrome mutations in type IV tropocollagen alter molecular structure and nanomechanical properties. *J. Struct. Biol.* **2009**, *168*, 503–510.
- (6) Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *P. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 4273–4278.
- (7) Sasaki, N.; Odajima, S. Stress-strain curve and Young's modulus of a collagen molecule as determined by the X-ray diffraction technique. *J. Biomech.* **1996**, *29*, 655–658.
- (8) Bozec, L.; Horton, M. Topography and mechanical properties of single molecules of type I collagen using atomic force microscopy. *Biophys. J.* **2005**, *88*, 4223–4231.
- (9) Sun, Y.; Luo, Z.; Fertala, A.; An, K. Direct quantification of the flexibility of type I collagen monomer. *Biochem. Biophys. Res. Commun.* **2002**, *295*, 382–386.
- (10) Fratzl, P.; Weinkamer, R. Nature's hierarchical materials. *Prog. Mater. Sci.* **2007**, *52*, 1263–1334.
- (11) Persikov, A. V.; Ramshaw, J. A.; Brodsky, B. Collagen model peptides: Sequence dependence of triple-helix stability. *Biopolymers* **2000**, *55*, 436–450.
- (12) Vesentini, S.; Fitie, C. F. C.; Montevecchi, F. M.; Redaelli, A. Molecular assessment of the elastic properties of collagen-like homotrimer sequences. *Biomech. Model Mechanobiol.* **2005**, *3*, 224–234.
- (13) Lorenzo, A. C.; Caffarena, E. R. Elastic properties, Young's modulus determination and structural stability of the tropocollagen molecule: a computational study by steered molecular dynamics. *J. Biomech.* **2005**, *38*, 1527–1533.
- (14) Buehler, M. J. Atomistic and continuum modeling of mechanical properties of collagen: Elasticity, fracture and self-assembly. *J. Mater. Res.* **2006**, *21*, 1947–1961.
- (15) Buehler, M. J. Nature designs tough collagen: Explaining the nanostructure of collagen fibrils. *P. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12285–12290.
- (16) Buehler, M. J.; Wong, S. Y. Entropic elasticity controls nanomechanics of single tropocollagen molecules. *Biophys. J.* **2007**, *93*, 37–43.
- (17) Gautieri, A.; Vesentini, S.; Montevecchi, F. M.; Redaelli, A. Mechanical properties of physiological and pathological models of collagen peptides investigated via steered molecular dynamics simulations. *J. Biomech.* **2008**, *41*, 3073–3077, DOI: 10.1016/j.jbiomech.2008.06.028.
- (18) Ayton, G. S.; Noid, W. G.; Voth, G. A. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.
- (19) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, *128*.
- (20) Wang, Y.; Noid, W. G.; Liu, P.; Voth, G. A. Effective force coarse-graining. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2002–15.
- (21) Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.
- (22) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (23) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (24) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hunenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
- (25) Chetwynd, A. P.; Scott, K. A.; Mokrab, Y.; Sansom, M. S. CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations. *Mol. Membr. Biol.* **2008**, *25*, 662–9.
- (26) Sansom, M. S.; Scott, K. A.; Bond, P. J. Coarse-grained simulation: a high-throughput computational approach to membrane proteins. *Biochem. Soc. Trans.* **2008**, *36*, 27–32.

- (27) Treptow, W.; Marrink, S. J.; Tarek, M. Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. *J. Phys. Chem. B* **2008**, *112*, 3277–82.
- (28) Risselada, H. J.; Marrink, S. J. Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2056–67.
- (29) Berendsen, H. J. C.; Vandespoel, D.; Vandrunen, R. Gromacs - a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (30) Vandrunen, R.; Vandespoel, D.; Berendsen, H. J. C. Gromacs - a software package and a parallel computer for molecular-dynamics. *Abstr. Pap. Am. Chem. Soc.* **1995**, *209*, 49.
- (31) Gautieri, A.; Buehler, M. J.; Redaelli, A. Deformation rate controls elasticity and unfolding pathway of single tropocollagen molecules. *J. Mech. Behav. Biomed. Mater.* **2009**, *2*, 130–137.
- (32) Rainey, J.; Goh, M. An interactive triple-helical collagen builder. *Bioinformatics* **2004**, *20*, 2458–2459.
- (33) Black, S. D.; Mould, D. R. Development of hydrophobicity parameters to analyze proteins which bear posttranslational or cotranslational modifications. *Anal. Biochem.* **1991**, *193*, 72–82.
- (34) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (35) Chu, J. W.; Voth, G. A. Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. *Biophys. J.* **2006**, *90*, 1572–1582.
- (36) Hofmann, H.; Voss, T.; Kuhn, K.; Engel, J. Localization Of Flexible Sites In Thread-Like Molecules From Electron-Micrographs - Comparison Of Interstitial, Basement-Membrane And Intima Collagens. *J. Mol. Biol.* **1984**, *172*, 325–343.
- (37) Utiyama, H.; Sakato, K.; Ikehara, K.; Setsuiye, T.; Kurata, M. Flexibility Of Tropocollagen From Sedimentation And Viscosity. *Biopolymers* **1973**, *12*, 53–64.
- (38) Saito, T.; Iso, N.; Mizuno, H.; Onda, N.; Yamato, H.; Odashima, H. Semi-Flexibility Of Collagens In Solution. *Biopolymers.* **1982**, *21*, 715–728.
- (39) Nestler, F. H. M.; Hvidt, S.; Ferry, J. D.; Veis, A. Flexibility Of Collagen Determined From Dilute-Solution Viscoelastic Measurements. *Biopolymers* **1983**, *22*, 1747–1758.
- (40) Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. Entropic Elasticity Of Lambda-Phage Dna. *Science.* **1994**, *265*, 1599–1600.
- (41) Pampaloni, F.; Lattanzi, G.; Jonas, A.; Surrey, T.; Frey, E.; Florin, E. L. Thermal fluctuations of grafted microtubules provide evidence of a length-dependent persistence length. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10248–10253.
- (42) Mogilner, A.; Oster, G. Cell motility driven by actin polymerization. *Biophys. J.* **1996**, *71*, 3030–3045.
- (43) Buehler, M. J. Atomistic modeling of elasticity, plasticity and fracture of protein crystals. *J. Comput. Theor. Nanosci.* **2006**, *3*, 670–683.
- (44) Sasaki, N.; Odajima, S. Elongation mechanism of collagen fibrils and force-strain relations of tendon at each level of structural hierarchy. *J. Biomech.* **1996**, *29*, 1131–1136.
- (45) Cusack, S.; Miller, A. Determination Of The Elastic-Constants Of Collagen By Brillouin Light-Scattering. *J. Mol. Biol.* **1979**, *135*, 39–51.
- (46) Higuchi, H.; Nakauchi, Y.; Maruyama, K.; Fujime, S. Characterization of beta-connectin (titin-2) from striated-muscle by dynamic light-scattering. *Biophys. J.* **1993**, *65*, 1906–1915.
- (47) Svoboda, K.; Schmidt, C. F.; Branton, D.; Block, S. M. Conformation and elasticity of the isolated red-blood-cell membrane skeleton. *Biophys. J.* **1992**, *63*, 784–793.
- (48) Sun, Y. L.; Luo, Z. P.; Fertala, A.; An, K. N. Direct quantification of the flexibility of type I collagen monomer. *Biochem. Biophys. Res. Commun.* **2002**, *295*, 382–386.
- (49) Saito, T.; Iso, N.; Mizuno, H.; Onda, N.; Yamato, H.; Odashima, H. Semiflexibility of collagens in solution. *Biopolymers.* **1982**, *21*, 715–28.
- (50) McCauley, M. J.; Williams, M. C. Mechanisms of DNA binding determined in optical tweezers experiments. *Biopolymers.* **2007**, *85*, 154–168.
- (51) Mucke, N.; Kreplak, L.; Kirmse, R.; Wedig, T.; Herrmann, H.; Aebi, U.; Langowski, J. Assessing the flexibility of intermediate filaments by atomic force microscopy. *J. Mol. Biol.* **2004**, *335*, 1241–1250.
- (52) Kikumoto, M.; Kurachi, M.; Tosa, V.; Tashiro, H. Flexural rigidity of individual Microtubules measured by a buckling force with optical traps. *Biophys. J.* **2006**, *90*, 1687–1696.